

LIVRE BLANC

DATA QUALITY

À horizon 2027

Pourquoi les données dégradées coûtent cher aux entreprises
et comment mettre en œuvre un DQM rentable



Data Quality

À horizon 2027

Pourquoi les données dégradées coûtent cher aux entreprises
et comment mettre en œuvre un DQM rentable

Ce document comprend :

	À propos de ce livre blanc	3
01	Diagnostic Data Quality : pourquoi 2026 a changé la donne	4
02	Comprendre concrètement la qualité des données	5
03	Quel est le prix d'une mauvaise qualité des données ?	8
04	Preuves par l'exemple avec 5 cas d'usages	10
05	Le DQM comme dénominateur commun entre stratégie, marketing et production	13
06	Anticiper ses besoins en data quality pour 2027	15
07	Feuille de route DQM : par où commencer ?	17
08	Pourquoi consulter IDAIA ?	19
	Sources	21

À propos de ce livre blanc

La fiabilité des données conditionne directement la justesse des décisions stratégiques, la performance commerciale et de plus en plus, la visibilité d'une entreprise face aux intelligences artificielles qui filtrent une part croissante de la recherche d'information. Ce document dresse un diagnostic chiffré de ce que coûte réellement une donnée dégradée, avant d'expliquer schémas à l'appui, les mécanismes concrets qui permettent de la fiabiliser. Cinq cas d'entreprises françaises et internationales, de la finance à l'industrie en passant par le marketing direct, illustrent comment des organisations très différentes ont transformé ce chantier en résultats mesurables. Une même conviction traverse ces exemples : la qualité des données n'est ni un projet informatique isolé ni un luxe réservé aux grandes structures. Elle reflète un langage commun entre la production, le pilotage financier et l'activation commerciale. À l'approche de 2027, entre accélération réglementaire française et bascule mondiale vers l'IA agentique, ce document propose enfin une feuille de route en 3 phases, pensée pour produire des résultats tangibles sans avoir à couvrir l'intégralité des données du SI de l'entreprise.



Diagnostic Data Quality : pourquoi 2026 a changé la donne

En 2026, la donnée conditionne directement la capacité d'une entreprise à décider, à automatiser et, de plus en plus, à exister aux yeux des intelligences artificielles qui filtrent désormais une part croissante de la recherche d'informations métier et commerciales. Or cette dépendance rend chaque défaut de qualité proportionnellement plus coûteux.

Le constat chiffré est sans ambiguïté, à l'échelle mondiale d'abord. Gartner estime que l'impact financier moyen d'une mauvaise qualité des données atteint 12,9 millions de dollars par an pour une organisation. Aux États-Unis spécifiquement, une analyse du cabinet Doubletrack publiée en février 2026 évalue à 617 milliards de dollars le coût annuel cumulé de ces données dégradées pour l'ensemble de l'économie américaine, l'équivalent d'environ 2 points de produit intérieur brut. Le secteur de l'information, hautement dépendant de la fiabilité de ses données, y perd à lui seul plus de 12 000 dollars par salarié chaque année. Ces montants ne sont pas des abstractions statistiques. Ce sont des heures de correction manuelle, des campagnes mal ciblées, des décisions prises sur des bases fausses.

L'exemple d'Unity Technologies illustre ce risque de façon assez nette. En 2022, l'algorithme de placement publicitaire de l'entreprise a ingéré des données biaisées transmises par un client majeur. Les performances du système se sont dégradées jusqu'à nécessiter une reconstruction complète, avec pour conséquence directe une chute de 37% du cours de l'action et un impact estimé à 110 millions de dollars. Un incident de qualité de données, pas une panne technique, pas une faille de sécurité, mais l'utilisation d'informations d'entrée erronées.

L'intelligence artificielle change l'échelle du problème

Ce qui a distingué 2026 des années précédentes, c'est l'accélération de l'IA dans les processus d'entreprise. Les dépenses mondiales en intelligence artificielle devraient dépasser 2 000 milliards de dollars cette année, en hausse de 37% sur un an selon Gartner. Chaque euro investi dans un système qui apprend sur des données médiocres amplifie mécaniquement l'erreur au lieu de la corriger. C'est le principe bien connu du garbage in, garbage out, désormais appliqué à grande échelle. Au niveau mondial toujours, Gartner prévoyait dès 2024 qu'au moins 30% des projets d'intelligence artificielle générative seraient abandonnés après la preuve de concept d'ici fin 2025, en raison notamment d'une qualité de données insuffisante. Un directeur des opérations sur deux, selon l'IBM Institute for Business Value, place désormais la qualité des données au sommet de ses priorités, devant les nouvelles plateformes d'IA elles-mêmes.

Un défi particulièrement français

Les entreprises françaises ne font pas exception. L'Observatoire de la maturité data et IA des entreprises, qui a interrogé près de 300 organisations françaises fin 2025, révèle un écart frappant : 88% des dirigeants jugent la donnée déterminante pour leur performance, mais seuls 19% d'entre eux s'estiment réellement capables de l'exploiter efficacement. Plus d'un répondant sur deux pilote encore son activité principalement via Excel ou Google Sheets. Disposer de données en volume est devenu banal. Savoir en tirer parti avec confiance et à grande échelle reste l'exception, y compris en France.

C'est précisément ce basculement d'un problème technique contenu à un enjeu de compétitivité globale que ce document se propose d'éclairer.

Comprendre concrètement la qualité des données

La qualité des données constitue une discipline à part entière, avec ses normes internationales, ses méthodes de mesure et ses écoles de pensée, au même titre que la qualité industrielle ou la conformité réglementaire. Elle mérite d'être traitée avec la même rigueur. Son principe fondateur, une fois mis en perspective, tient néanmoins en une phrase : une donnée est de qualité lorsqu'elle permet réellement de faire ce qu'on attend d'elle, au moment où on en a besoin. La norme internationale ISO 25012, qui fait référence en la matière, la définit comme le degré auquel un ensemble de données possède les caractéristiques nécessaires pour satisfaire des besoins exprimés dans un contexte d'usage donné. Autrement dit, la même donnée peut être excellente pour un usage et inutilisable pour un autre.

Prenons un exemple qui parlera à quiconque manipule une base de prospection ou un fichier client. Une fiche contact indique un nom, une adresse email et un numéro de téléphone portable, saisis voilà 3 ans. Cette fiche est-elle de qualité ? La réponse dépend entièrement de ce qu'on en fait.

Les dimensions qui font la différence

Les professionnels de la donnée décomposent traditionnellement la qualité en plusieurs dimensions, chacune répondant à une question précise sur notre fiche contact.

- **L'exactitude** interroge la conformité au réel : ce numéro de téléphone appartient-il toujours à la bonne personne ?
- **La complétude** vérifie qu'aucune information nécessaire ne manque : dispose-t-on du code postal indispensable à une campagne géolocalisée ?
- **La fraîcheur**, ou actualité, questionne la pertinence temporelle : une adresse relevée il y a 3 ans reflète-t-elle encore la réalité du lieu d'habitation ?
- **L'unicité** s'assure qu'un même contact n'existe pas en plusieurs exemplaires, sous des orthographes légèrement différentes, faussant ainsi le décompte réel de la base.
- **La cohérence** vérifie l'harmonie entre systèmes : le même client porte-t-il la même civilité dans le CRM et dans l'outil de facturation ?
- **La validité** contrôle enfin la conformité aux règles définies : un email respecte-t-il un format exploitable, un code SIREN comporte-t-il bien 14 chiffres ?

Une seule fiche défaillante sur une de ces dimensions reste anecdotique. Multipliée par plusieurs milliers de contacts, elle devient un problème structurel qui fausse silencieusement chaque campagne, chaque prévision, chaque décision qui s'appuie sur cette base.

Une qualité qui se définit toujours par l'usage

Le Laboratoire national de métrologie britannique, référence académique du secteur, insiste sur ce principe souvent négligé. La qualité d'une donnée ne se juge donc jamais dans l'absolu, elle se juge par rapport à l'usage qu'on veut en faire. Une même dimension change de sens selon le contexte. La traçabilité d'un email, par exemple, ne recouvre pas la même exigence pour une campagne marketing classique, où l'on souhaite simplement savoir d'où vient l'adresse, que pour un dossier de conformité réglementaire, où il faut pouvoir démontrer l'origine exacte du consentement et la date précise de sa collecte.

Cette nuance change tout dans la pratique. Elle signifie qu'il n'existe pas de note universelle de qualité des données, mais des niveaux d'exigence différents selon que l'on prépare une campagne d'activation commerciale, une analyse stratégique ou un audit de conformité. C'est cette exigence variable, et la manière d'y répondre méthodiquement, que nous allons maintenant traduire en coûts réels pour l'entreprise.

Les mécanismes opérationnels se jouent derrière chaque donnée fiable

Comprendre les dimensions de la qualité ne suffit évidemment pas à savoir comment on l'obtient concrètement. Derrière une base de données jugée fiable se cache un enchaînement de mécanismes précis, que les équipes data désignent sous le nom de chaîne de traitement.

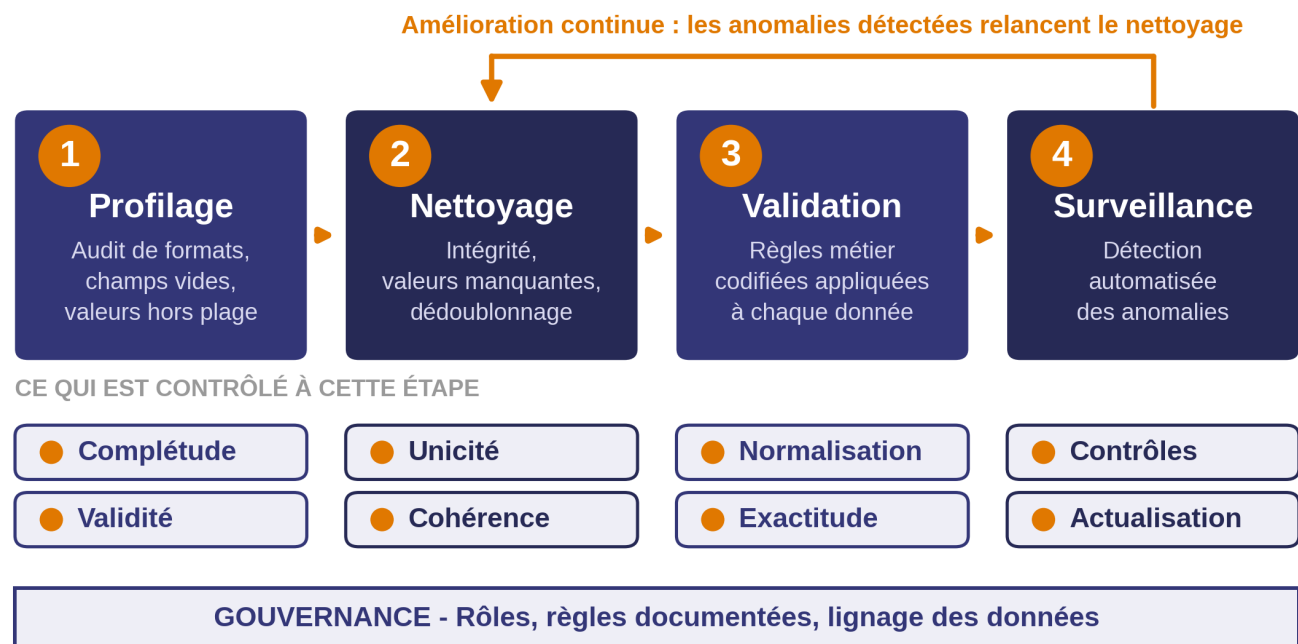
Le profilage, première étape, consiste à interroger systématiquement l'existant avant toute intervention. Il s'agit d'identifier les formats de données qui coexistent réellement dans une base, la proportion de champs obligatoires restés vides, ou encore les valeurs qui sortent des plages attendues. Cette analyse produit un diagnostic chiffré, un point de référence à partir duquel toute amélioration pourra ensuite être mesurée.

Le nettoyage, qui suit logiquement, regroupe les opérations de normalisation des formats, de traitement des valeurs manquantes et surtout de dédoublonnage. Cette dernière tâche, en apparence anodine, repose en réalité sur des techniques de correspondance approximative capables de repérer qu'une même entreprise, orthographiée différemment d'une saisie à l'autre, désigne bien le même enregistrement.

La validation applique ensuite à chaque donnée un ensemble de règles métier codifiées. Un moteur de règles vérifie qu'un numéro SIREN comporte bien 14 chiffres, qu'une date de naissance ne se situe pas dans le futur, qu'un code postal correspond effectivement à la ville renseignée. Ces contrôles, une fois automatisés, interceptent l'essentiel des anomalies avant qu'elles n'atteignent les systèmes de décision.

Le cycle opérationnel de la qualité des données

Quatre étapes dans une boucle de correction continue sous gouvernance



Source : conception IDAIA, d'après les référentiels DAMA et ISO 8000 et le NPL Data Quality Good Practice Guide (mars 2026)

La surveillance continue prend le relais dans la durée. Une base de données, aussi propre soit elle au moment de sa livraison, se dégrade naturellement avec le temps. Des coordonnées deviennent obsolètes faute d'actualisation, des doublons réapparaissent après une fusion de systèmes, des dérives s'installent silencieusement dans un format. Des outils d'observabilité, de plus en plus soutenus par l'apprentissage automatique, détectent ces anomalies quasiment en temps réel, sans attendre un audit ponctuel.

La gouvernance, enfin, ne constitue pas une cinquième étape qui suivrait les précédentes, mais le cadre qui les rend pérennes. Elle attribue des responsabilités claires, un responsable de domaine métier qui porte la vision d'ensemble d'une famille de données, un référent opérationnel qui en assure le suivi quotidien, documente les règles dans un catalogue partagé, et trace la provenance de chaque donnée à travers ce que les spécialistes appellent son lignage, soit l'historique complet des transformations qu'elle a subies depuis sa création.

Quel est le prix d'une mauvaise qualité des données ?

La particularité de ce sujet, qui explique en partie pourquoi il reste sous-investi dans tant d'organisations, tient à une asymétrie de perception. Les moyens consacrés à la qualité des données, un poste de data steward, un outil de nettoyage, du temps d'équipe, sont immédiatement visibles dans un budget. Le coût de la non-qualité, lui, se dilue dans des centaines de micro-décisions quotidiennes et n'apparaît jamais comme une ligne isolée sur un compte de résultat. Cette invisibilité comptable ne rend pas le problème moins réel, mais simplement plus difficile à faire remonter jusqu'aux instances qui arbitrent les investissements.

Quatre familles de risques issues d'une même origine

Chiffre d'affaires

Le groupe de travail Qualité des données constitué par DAMA France et l'ISACA AFAI estime qu'une insuffisance de qualité peut, selon les organisations, amputer le chiffre d'affaires de 15 à 25%, à travers des décisions prises sur des bases erronées, des campagnes mal ciblées ou des opportunités commerciales tout simplement invisibles faute de données exploitables.

Réputation de l'entreprise

Une base clients de mauvaise qualité mobilisée dans une campagne de communication peut ternir durablement la relation de confiance avec des parties prenantes, qu'il s'agisse de particuliers sollicités par erreur ou de partenaires professionnels dont les informations n'ont pas été mises à jour.

Coûts directs

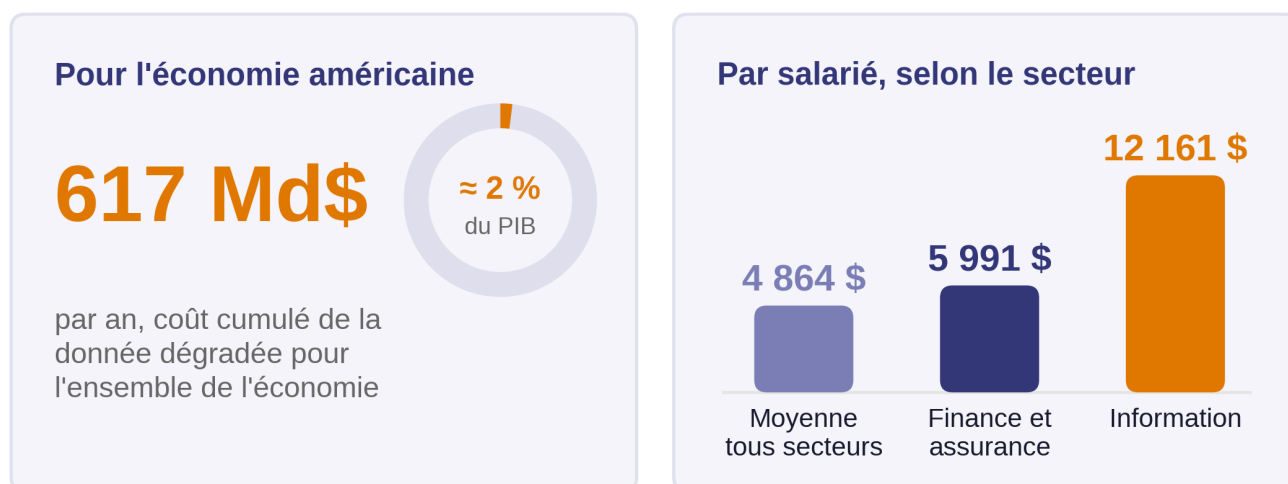
Chaque anomalie détectée en aval déclenche une chaîne de corrections manuelles, de vérifications et de retraitements, dont le coût réel dépasse presque toujours l'estimation initiale, précisément parce qu'il se répartit sur de nombreuses équipes plutôt que de se concentrer sur un seul centre de coût identifiable.

Conformité réglementaire

Pour une entreprise soumise aux obligations de lutte contre le blanchiment, à la vérification d'identité ou à la fiabilisation des coordonnées bancaires, une donnée défaillante n'est plus seulement un problème d'efficacité opérationnelle, elle devient un motif de sanction. Une enquête menée à l'échelle mondiale par Forrester révèle que plus d'un quart des professionnels de la donnée estiment que la mauvaise qualité coûte plus de 5 millions de dollars par an à leur organisation, et que 7% d'entre eux évoquent des pertes supérieures à 25 millions de dollars.

Aux États-Unis, la donnée dégradée coûte très cher

Coût annuel de la mauvaise qualité des données, à l'échelle de l'économie puis du salarié



Source : Doubletrack, « The Hidden Cost of Dirty Data », février 2026 (analyse sectorielle, économie américaine)

Pourquoi le retour sur investissement classique ne suffit pas

Face à ces montants, la tentation naturelle consiste à réclamer un calcul de retour sur investissement avant d'engager une démarche de mise en qualité. L'exercice se heurte pourtant à une difficulté structurelle. Si les coûts d'un projet de qualité des données sont généralement bien identifiables, une part significative des bénéfices attendus reste, par nature, spéculative. Reconstituer la confiance perdue d'un client, éviter une décision stratégique mal orientée ou prévenir une sanction réglementaire hypothétique ne se traduit pas aisément en euros vérifiables. Un calcul de retour sur investissement, dans ce contexte, présentera quasi systématiquement un coût précis et un bénéfice difficile à opposer.

Cela ne signifie pas qu'il faille renoncer à toute rigueur budgétaire, mais qu'il convient de circonscrire l'exercice à un périmètre suffisamment concret pour rester crédible. La méthode la plus robuste consiste à choisir un processus métier précis où l'usage de la donnée est parfaitement identifié, à relier chaque défaut de qualité observé à un indicateur mesurable (volume de dossiers incomplets, nombre de relances inutiles, taux de rejet...), puis à comparer le coût du traitement actuel à celui d'une donnée fiabilisée. Appliquée à un périmètre restreint et opposable, cette approche produit un chiffrage défendable devant une direction générale, là où une estimation globale et diffuse ne convaincra jamais personne.

C'est cette conviction née de cas concrets qui a permis à des organisations très différentes de franchir le pas. Les 5 exemples qui suivent en témoignent.

Preuves par l'exemple avec 5 cas d'usages

La théorie convainc rarement à elle seule un comité de direction. Ces organisations ont ceci de commun qu'elles ont toutes engagé une démarche de qualité des données à partir d'un problème concret, sans attendre un cadre parfait ni un budget illimité. Elles couvrent des secteurs proches des enjeux d'IDAIA, tels que la finance réglementée, le secteur public ou la performance commerciale, tout en démontrant qu'un principe de traitement applicable aux données industrielles peut aussi trouver un retentissement sur des données humaines.

BNP Paribas Asset Management

Avec plus de 3 000 collaborateurs répartis dans 37 pays, cette société de gestion d'actifs a créé dès janvier 2018 un Data Office pour répondre à des exigences réglementaires croissantes, notamment MIFID, le RGPD et l'accord bancaire BCBS 239. Sa doctrine tient en un sigle devenu culture d'entreprise : une donnée n'existe véritablement que si elle est définie, gouvernée et utile. Sans lancer de chantier lourd, l'équipe a organisé des ateliers courts réunissant producteurs et consommateurs de la donnée, afin de faire prendre conscience aux premiers de l'impact de leurs erreurs sur les seconds, et de recueillir en retour les difficultés concrètes rencontrées lors de la saisie. Cette méthode légère a permis de bâtir un plan d'action opposable sans mobiliser systématiquement les équipes informatiques, pour réduire les délais de résolution des anomalies et identifier les besoins de formation propres à chaque métier.

Caisse des Dépôts

Cet acteur financier public de 125 000 collaborateurs, organisé autour de 5 grands métiers allant des retraites au financement des entreprises, a engagé en 2018 un plan de transformation numérique avec l'ambition de devenir pleinement pilotée par la donnée. Le diagnostic initial révélait un manque d'information sur les données existantes, des responsabilités dispersées entre plusieurs acteurs et une qualité perçue plutôt que mesurée. La réponse a pris la forme d'un cadre de gouvernance structuré autour de 4 axes complémentaires : une organisation confiant à un responsable de domaine la gouvernance de ses données et à un référent opérationnel leur suivi quotidien, des normes communes applicables à toute l'entreprise, un processus continu d'amélioration partagé entre les parties prenantes, et un socle technologique associant catalogue de données et outil de contrôle qualité. Trois cas pilotes, menés en 3 mois sur un périmètre volontairement restreint de 150 types de données, ont suffi à démontrer la viabilité du dispositif avant son extension.

Un grand groupe industriel du BTP et des concessions

Confronté à la gestion de ses données de ressources humaines sur un effectif dépassant 150 000 salariés, ce groupe a entrepris une démarche de fiabilisation systématique de ses bases. Le taux d'erreur constaté sur ces données, initialement proche de 8%, a été ramené à 0,05% après la mise en place de contrôles et de processus de validation adaptés à cette échelle. Ce résultat illustre un principe souvent sous-estimé hors du monde industriel : une donnée de mauvaise qualité, comme une pièce défectueuse sur une chaîne de production, se corrige beaucoup plus facilement en amont, au moment de sa création, qu'une fois qu'elle a irrigué l'ensemble des systèmes en aval.

Un acteur du marketing direct et de la délivrabilité email

Spécialisé dans l'activation de campagnes emailing pour ses clients annonceurs, cet acteur a fait de la qualité de ses bases de contacts le socle même de sa promesse commerciale. En consolidant ses processus de validation d'adresses, de déduplication et de mise à jour régulière, il atteint aujourd'hui un taux de délivrabilité de 95%, un niveau qui conditionne directement le retour sur investissement de chaque campagne envoyée pour le compte de ses clients. L'exemple rappelle une évidence trop souvent oubliée par les équipes marketing : aucune créativité publicitaire ne compense une adresse email invalide ou une fiche CRM en plusieurs exemplaires.

Un groupe hôtelier et de loisirs

Ce groupe de près de 6 000 collaborateurs, réparti sur une cinquantaine d'établissements entre hébergement, restauration et divertissement, a vu ses ambitions de digitalisation freinées par une base clients construite sur 2 métiers qui ne collectaient pas les mêmes informations d'identification. Les doublons se multipliaient, les délais d'alimentation entre systèmes dépassaient les attentes des équipes marketing, et le programme de fidélité fonctionnait sur une chaîne entièrement automatisée dont personne ne questionnait plus l'exactitude. La création d'un poste dédié au pilotage de la qualité, chargé d'instruire chaque anomalie avec les équipes concernées et de proposer les corrections nécessaires, a permis en un an de faire reconnaître une baisse mesurable des anomalies par les utilisateurs eux-mêmes de la donnée, preuve la plus convaincante qui soit dans une démarche de cette nature.



Activer 5 leviers adaptés pour des résultats concrets

Synthèse des 5 cas d'usage présentés en partie 4

Organisation	Problème de départ	Levier actionné	Résultat clé
BNP Paribas Asset Management <i>3 000 collaborateurs, 37 pays</i>	Exigences réglementaires croissantes (MiFID, RGPD, BCBS 239)	Ateliers courts réunissant producteurs et consommateurs de la donnée	Délais de résolution des anomalies réduits
Caisse des Dépôts <i>125 000 collaborateurs</i>	Qualité perçue plutôt que mesurée, responsabilités dispersées	Gouvernance en 4 axes : rôles, normes communes, amélioration continue, outils	Dispositif validé en 3 mois sur 150 données pilotes avant extension
Groupe industriel BTP et concessions <i>Plus de 150 000 salariés</i>	Données RH peu fiables, taux d'erreur proche de 8 %	Contrôles et validation adaptés à grande échelle	Taux d'erreur ramené de 8 % à 0,05 %
Acteur du marketing direct et délivrabilité <i>Emailing pour annonceurs</i>	Qualité des bases de contacts au cœur de la promesse commerciale	Validation d'adresses, déduplication, mise à jour régulière	Taux de délivrabilité de 95 % atteint
Groupe hôtelier et de loisirs <i>Près de 6 000 collaborateurs, une cinquantaine d'établissements</i>	Doublons et deux métiers aux identifiants non alignés	Poste dédié au pilotage de la qualité des données	Baisse mesurable des anomalies reconnue par les utilisateurs

Source : DAMA France et ISACA AFAl, cahier pratique n°1 (2021) ; cas publics documentés de fiabilisation de données RH et de délivrabilité email

Au-delà de leurs différences sectorielles, ces 5 organisations partagent une même leçon : la qualité des données s'améliore par le biais d'une méthode outillée, portée par des rôles clairement identifiés et mesurée dans la durée. C'est cette méthode que nous allons maintenant détailler.

Le DQM comme dénominateur commun entre stratégie, marketing et production

Une entreprise qui traite la qualité des données comme un sujet purement informatique passe à côté de l'essentiel. La donnée irrigue simultanément 3 strates de l'organisation qui, en apparence, n'ont que peu en commun : les opérations qui font tourner la production, la direction qui pilote la performance financière, et le marketing/commerce qui doit convertir des prospects en clients. Ces 3 strates partagent pourtant une même dépendance. Une donnée corrompue à la source ne reste jamais confinée à son service d'origine, elle se propage et fausse, en cascade, chacun des usages qui en dépendent.

C'est cette transversalité qui fait de la gestion de la qualité des données un langage commun entre des fonctions qui, trop souvent, s'ignorent. Un défaut dans une donnée de production remonte jusqu'au tableau de bord du directeur financier. Une base commerciale mal tenue prive la direction générale d'une vision fiable de son portefeuille clients. Une fiche produit incomplète compromet à la fois la chaîne logistique et la campagne publicitaire censée la vendre. Traiter ces 3 dimensions séparément revient à soigner un symptôme sans jamais s'attaquer à sa cause commune.

5.1 Données métier : le socle des opérations et de la production

Dans l'industrie, la donnée ne décrit pas une abstraction commerciale, elle décrit une réalité physique, une référence produit, une nomenclature, un niveau de stock, une capacité machine. Une erreur sur ces données entraîne plus qu'une simple gêne statistique, elle se traduit directement par une ligne de production arrêtée, une commande fournisseur erronée ou une rupture de stock non anticipée. L'exemple industriel évoqué plus haut, où un taux d'erreur sur des données de plusieurs centaines de milliers de collaborateurs a été ramené de 8% à 0,05%, illustre à quel point cette discipline s'apparente en termes d'exigence au contrôle qualité que ces mêmes entreprises appliquent déjà à leurs produits physiques. Personne n'accepterait qu'une pièce mécanique présente un taux de défaut de 8% en sortie de chaîne. La donnée qui pilote cette chaîne mérite la même vigilance.

5.2 Données de pilotage : décider juste, du chiffre d'affaires à la rentabilité

À l'échelon de la direction générale et de la direction financière, la donnée prend une autre forme : indicateurs consolidés, prévisions, tableaux de bord de rentabilité. Ces chiffres ont ceci de trompeur qu'ils inspirent une confiance instinctive du seul fait de leur présentation soignée, quelle que soit la fiabilité réelle des données qui les alimentent en amont. Une prévision de chiffre d'affaires bâtie sur un référentiel client incomplet, un calcul de rentabilité par segment reposant sur des coûts mal ventilés, une consolidation financière agrégeant des données hétérogènes entre filiales : chacune de ces situations expose l'entreprise à des décisions stratégiques prises sur des bases fragiles, sans que personne, au moment de trancher, n'en ait pleinement conscience. C'est précisément dans ce contexte que la gouvernance des données, avec ses rôles et ses règles documentées, transforme un tableau de bord d'apparence fiable en outil robuste d'aide à la décision.

5.3 Données marketing et commerciales : investir sur les bons leviers et réussir l'activation

Le marketing occupe une position singulière dans cette chaîne puisque c'est souvent lui qui produit et exploite le plus grand volume de données ; c'est aussi lui qui en paie le prix le plus visible lorsqu'elles se dégradent, sous la forme de campagnes mal ciblées ou de budgets média gaspillés. Cette dépendance ne cesse de s'accroître avec la disparition progressive des cookies tiers, qui pousse les annonceurs à réinvestir massivement dans leurs données propriétaires pour alimenter des algorithmes publicitaires de plus en plus automatisés.

Les forces de vente dépendent tout autant de cette rigueur, même si leur rapport à la donnée reste souvent moins visible que celui du marketing. Une base de prospection mal qualifiée conduit un commercial à démarcher un interlocuteur qui a changé de poste depuis des mois, ou une entreprise qui ne correspond plus au profil recherché, avec pour conséquence directe un temps de prospection gaspillé et un discours commercial mal calibré dès le premier échange. L'historique de suivi conservé dans le CRM pose une exigence similaire ; un pipeline commercial construit sur des fiches dupliquées ou des statuts d'opportunité obsolètes fausse à la fois les prévisions de vente et la priorisation des relances, 2 leviers pourtant décisifs pour la performance d'une équipe commerciale. Une base de prospection fraîche, correctement segmentée et débarrassée de ses doublons ne relève donc pas du confort, elle conditionne directement le taux de transformation de chaque effort commercial engagé en amont.

Un phénomène plus récent vient renforcer cet enjeu. Les intelligences artificielles génératives, lorsqu'elles répondent à une requête en ligne, ne citent que les sources qu'elles jugent fiables et documentées. Une entreprise dont les données de marché, les études internes ou les cas clients restent approximatifs ou non structurés se prive mécaniquement d'une visibilité croissante dans ces réponses, quand un concurrent capable de produire une donnée propriétaire rigoureuse s'y trouve mentionné en référence. Des variantes de ce principe s'appliquent notamment au géomarketing, où la pertinence d'un ciblage géographique dépend de l'exactitude des adresses et des zones de chalandise mobilisées, et au tracking utilisateur, où une donnée comportementale mal qualifiée fausse chaque optimisation ultérieure. Le marketing performant de demain ne se distingue plus par sa créativité seule, il se définit à travers la qualité de la matière première par laquelle l'ingéniosité s'exerce.

Anticiper ses besoins en data quality pour 2027

Il s'agit désormais de se projeter à horizon 2027, en visualisant la convergence simultanée d'une accélération réglementaire propre à la France et d'une bascule technologique mondiale portée par l'intelligence artificielle. Les entreprises qui attendront encore pour s'y préparer découvriront sans doute que la qualité des données n'était pas le premier sujet à différer indéfiniment.

Un calendrier réglementaire français qui ne laisse plus de marge

Plusieurs échéances se resserrent sur cette période en France. L'obligation de facturation électronique s'étendra aux petites et moyennes entreprises en septembre 2027, imposant la fiabilisation des données fournisseurs et clients bien avant cette date pour les organisations encore hors chantier. La sixième directive européenne de lutte contre le blanchiment doit être transposée en droit français en juillet 2027, accompagnée par l'ouverture d'une agence européenne dédiée pleinement opérationnelle dès janvier 2028. Ensemble, ces échéances dessinent une même exigence, des référentiels de données propres, documentés et vérifiables, sans lesquels aucune de ces obligations ne peut être satisfaite dans les délais.

L'intelligence artificielle rebat les cartes plus vite que prévu

À l'échelle mondiale, Gartner livre un diagnostic sans détour sur ce qui attend les organisations d'ici 2027. Les entreprises qui structurent correctement la sémantique de leurs données, c'est-à-dire qui documentent précisément ce que chaque donnée représente et comment elle s'articule avec les autres, pourront améliorer jusqu'à 80% la précision de leurs modèles d'intelligence artificielle générative et réduire jusqu'à 60% les coûts associés. Dans le même temps, la moitié des décisions d'entreprise devrait être augmentée ou automatisée par des agents IA. Cette promesse s'accompagne toutefois d'un revers que peu d'organisations anticipent. Gartner prévoit que 80% des initiatives de gouvernance des données échoueront d'ici 2027, faute d'un portage suffisamment fort au sein des directions, et que plus de 40% des projets d'IA agentique seront purement et simplement annulés sur la même période, pour des raisons de coûts incontrôlés ou de valeur business jamais démontrée.

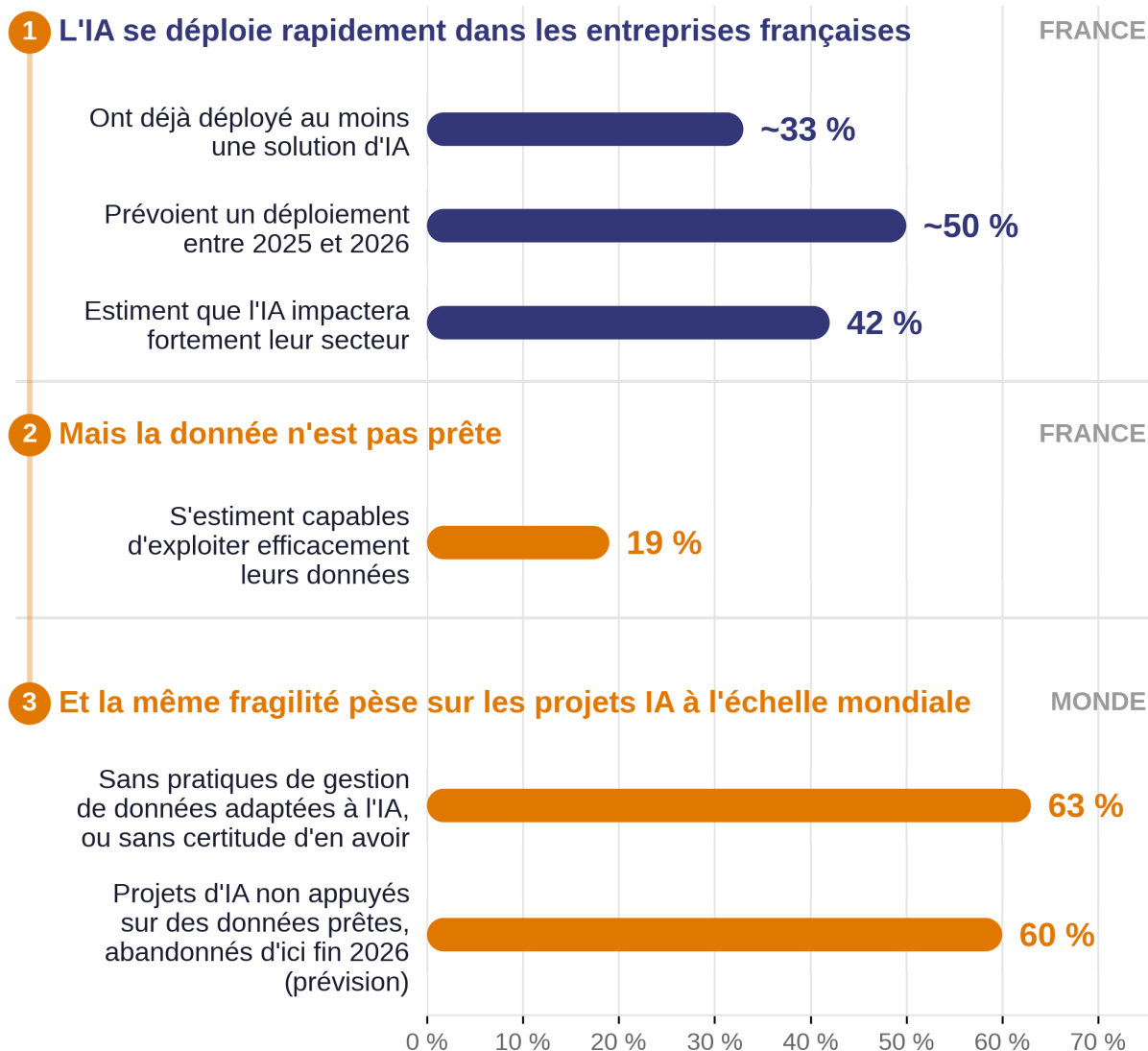
Le point commun entre ces réussites et ces échecs tient presque toujours à la même variable, la qualité et la structuration des données sur lesquelles ces systèmes s'appuient. Une intelligence artificielle, aussi sophistiquée soit-elle, ne compense jamais une donnée d'entrée défailante, elle l'amplifie.

Les entreprises françaises accélèrent sans mesurer le socle nécessaire

L'Observatoire de la maturité data et IA des entreprises, qui a interrogé près de 300 organisations françaises fin 2025, confirme cette dynamique sur le terrain. Près d'une entreprise sur trois a déjà déployé au moins une solution d'intelligence artificielle, et une sur deux prévoit un déploiement entre 2025 et 2026. Parmi les dirigeants interrogés, 74% estiment que l'IA impactera leur secteur, dont 42% l'estiment fortement.

L'IA se déploie plus vite que la donnée qui l'alimente

Déploiement de l'IA et préparation des données, en France puis dans le monde



Sources : Observatoire de la maturité data et IA des entreprises / Datasulting, résultats 2025 publiés en février 2026 (France) ; Gartner, février 2025 (enquête de juillet 2024 auprès de 1 203 responsables de la gestion des données, prévision jusqu'en 2026 / monde)

Cet écart entre l'ambition et la capacité réelle constitue, presque mot pour mot, la définition d'un risque non maîtrisé. Une entreprise qui déploie un agent conversationnel ou un outil de scoring prédictif sur des données de prospection dupliquées ou obsolètes ne fait qu'automatiser ses erreurs à plus grande échelle et à plus grande vitesse.

Se préparer pour 2027 ne consiste donc pas à anticiper une technologie précise, dont la forme exacte reste d'ailleurs difficile à prévoir. Il s'agit de construire dès maintenant le socle de données fiable, gouverné et documenté, sans lequel aucune des promesses de l'intelligence artificielle ni aucune des obligations réglementaires à venir ne saurait être tenue.

Feuille de route DQM : par où commencer ?

Face à l'ampleur du sujet, la tentation est grande de vouloir tout corriger en même temps, ce qui aboutit souvent à un chantier qui s'enlise avant de produire le moindre résultat visible. Les organisations qui réussissent leur démarche de mise en qualité des données procèdent différemment : elles avancent par paliers courts, sur un périmètre restreint et mesurable, avant d'étendre progressivement le dispositif. La méthode qui suit s'inspire directement des retours d'expérience présentés en partie 4, où des cycles de 3 mois sur une centaine de données ciblées ont suffi à convaincre des directions entières.

Phase 1 : Diagnostiquer avant d'agir

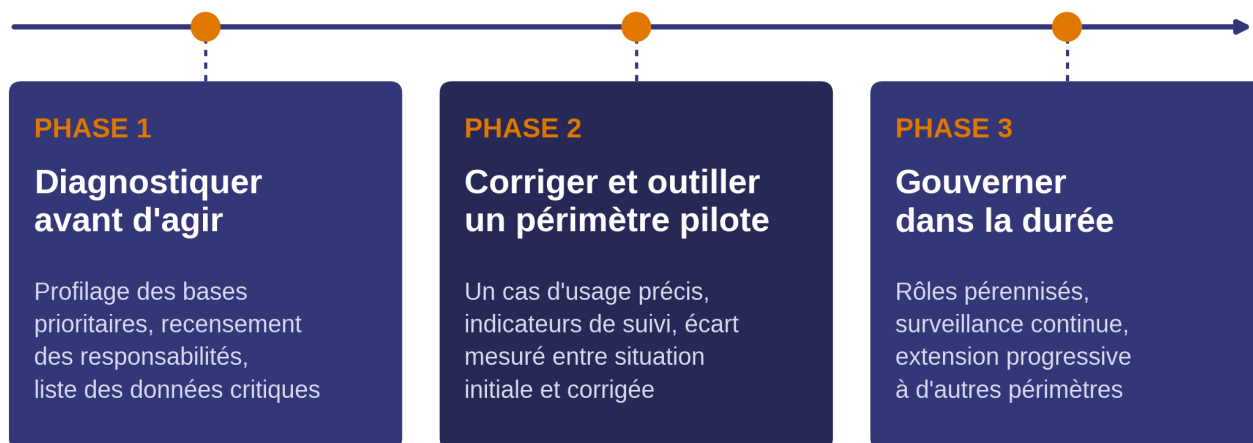
Cette première étape privilégie un état des lieux honnête, condition préalable à toute correction efficace. Elle passe par un profilage des bases prioritaires, celles qui alimentent la prospection commerciale, le reporting financier ou la production, selon les enjeux propres à chaque organisation, ainsi que par un recensement des responsabilités existantes, souvent plus dispersées qu'on ne l'imagine. C'est également à ce stade que se dessine la liste des données dites critiques, celles dont un défaut de qualité aurait l'impact le plus lourd sur le chiffre d'affaires, la conformité ou la relation client, en vue de concentrer l'effort là où il rapportera le plus.

Phase 2 : Corriger et outiller un périmètre pilote

Cette phase se concentre volontairement sur un cas d'usage précis et encadré : une base de prospection, un référentiel produit, un fichier de conformité réglementaire. Il ne s'agit pas de traiter l'ensemble du système d'information d'un coup, mais d'y déployer les mécanismes détaillés en partie 2, nettoyage, dédoublonnage, règles de validation, sur ce périmètre restreint, avec des indicateurs de suivi définis dès le départ pour démontrer, chiffres à l'appui, l'écart entre la situation initiale et la situation corrigée. Ce chiffrage circonscrit et vérifiable permettra ensuite de justifier l'extension du dispositif, là où une estimation globale et invérifiable convainc rarement, à juste titre.

Feuille de route en 3 phases par paliers courts

De l'état des lieux à la gouvernance dans la durée



Source : conception IDAIA, méthode inspirée des cas Caisse des Dépôts et BNP Paribas Asset Management

Phase 3 : Gouverner dans la durée

Une base de données nettoyée une fois se dégrade inévitablement si personne n'en assure l'entretien. Cette dernière phase instaure les rôles qui pérennisent l'effort, avec un responsable de domaine qui porte la vision d'ensemble d'une famille de données, et un référent opérationnel qui en assure le suivi quotidien, documente les règles dans un catalogue partagé, et met en place une surveillance continue capable de détecter les nouvelles anomalies avant qu'elles ne se propagent. C'est également le moment d'étendre progressivement la méthode à d'autres périmètres, en s'appuyant sur la crédibilité acquise lors de la phase pilote.

Un principe transversal mérite d'être rappelé à ce stade : cette feuille de route s'applique aussi bien à une base de données de production qu'à un référentiel commercial ou une base marketing. Une équipe de vente qui souhaite fiabiliser son pipeline CRM suivra exactement la même logique en 3 phases qu'une direction industrielle qui cherche à sécuriser sa chaîne de fabrication. Elles s'appuieront naturellement sur des indicateurs de succès propres à leurs métiers respectifs, tout en appliquant une méthode globale identique.

Pourquoi consulter IDAIA ?

Ce document cherche à mettre en exergue une problématique simple à énoncer mais exigeante à intégrer dans sa roadmap. La qualité des données ne relève pas du seul confort technique ni d'une case à cocher réglementaire. Elle conditionne directement la performance des forces de vente, la fiabilité des orientations stratégiques, et structure un environnement numérique dans lequel l'IA flirte désormais en permanence avec l'information disponible. Reste une question légitime : par où commencer, et avec quel partenaire ?

Le marché français des services data se caractérise par une fragmentation nette ; d'un côté, des acteurs internationaux généralistes, solides sur le risque et la conformité, mais peu présents sur l'accompagnement opérationnel et la personnalisation. De l'autre, des plateformes récentes agiles sur la prospection en libre-service, mais sans la profondeur de traitement ni l'expérience nécessaires pour couvrir des enjeux réglementaires complexes. Entre ces 2 extrémités, peu d'acteurs français réunissent sous une même bannière la fiabilisation des données, leur mise en conformité, leur activation commerciale et leur sécurisation contre la fraude.

IDAIA occupe cette position en adossant ses compétences à 30 années d'expertise data, une intelligence artificielle souveraine développée et hébergée en interne, ainsi qu'une architecture organisée autour de 5 piliers complémentaires : la qualité et le traitement des données, leur mise en conformité réglementaire, la prospection commerciale, le data marketing et l'identité numérique. La largeur de cette gamme de services n'a rien d'anecdotique pour un dirigeant, un responsable data ou un directeur marketing, puisqu'elle signifie qu'une base de prospection fiabilisée, un scoring client prédictif, une campagne d'activation géomarketing et une vérification d'identité conforme relèvent d'un seul et même partenaire disposant d'une vision d'ensemble, et non de plusieurs prestataires dont les référentiels de données ne communiquent jamais entre eux.

Le partenaire unique d'une vision data intégrale

L'architecture IDAIA au service de la donnée fiable, conforme, activable et sécurisée



Source : positionnement et architecture IDAIA

Cette expertise se traduit notamment par la mise à disposition d'outils conçus pour répondre à des usages data précis. Le référentiel de typologie client IDAIA permet d'affiner un ciblage com-

mercial sur des critères réellement discriminants. La plateforme géomarketing Datadrive® facilite l'ajustement des zones de chalandise et place la donnée géographique au service des décisions d'implantation et d'activation locale. Enfin, la Data Factory, interrogeable en temps réel via une API (interface applicative), sécurise l'intégration de la donnée fiabilisée dans les systèmes déjà en place chez le client, sans rupture de chaîne.

En 2027, la qualité des données ne peut plus être reportée à un futur projet de transformation numérique. IDAIA accompagne ses clients dans cette démarche depuis plus de 3 décennies, avec la conviction qu'une donnée fiable reste, aujourd'hui plus que jamais, le meilleur investissement qu'une entreprise puisse faire sur elle-même.

Si vous souhaitez obtenir un diagnostic de qualité de vos données ou construire votre feuille de route DQM avec nos experts, n'hésitez pas à [nous contacter](#).

Rédigé par Guylaine Michelet, experte en performance digitale et data marketing chez IDAIA Group.



Sources

Retrouvez ci-dessous les sources externes mobilisées tout au long de ce document, pour chaque partie.

1. Diagnostic Data Quality	Gartner (via IBM Think, 2022 et 2024), IBM Institute for Business Value (2025), Doubletrack (février 2026), IBM Think « bad data » (cas Unity Technologies), Observatoire de la maturité data et IA des entreprises / Datasulting (résultats 2025, publiés février 2026).
2. Comprendre concrètement la qualité des données	ISO 25012 (DAMA et NPL), National Physical Laboratory, Data Quality Good Practice Guide (mars 2026), principe de contextualité de la qualité.
3. Quel est le prix d'une mauvaise qualité des données ?	DAMA France et ISACA AFAI, cahier pratique n°1 (2021) ; Forrester, cité par IBM Think (novembre 2025).
4. Preuves par l'exemple avec 5 cas d'usages	DAMA France et ISACA AFAI, cahier pratique n°1 (2021, cas BNP Paribas Asset Management, Caisse des Dépôts, groupe hôtelier et de loisirs) ; cas publics documentés de fiabilisation de données RH industrielles à grande échelle et de délivrabilité email.
5. Le DQM comme dénominateur commun entre stratégie, marketing et production	Cas industriel déjà cité en partie 4, dynamique de disparition des cookies tiers et enjeu de citation par les IA génératives.
6. Anticiper ses besoins en data quality pour 2027	Calendrier réglementaire français (facturation électronique, sixième directive anti-blanchiment européenne) ; Gartner, prévisions à horizon 2027 (communiqués de presse 2024-2026) ; Observatoire de la maturité data et IA des entreprises / Datasulting (résultats 2025, publiés février 2026).
7. Feuille de route DQM	Méthode inspirée des cas Caisse des Dépôts et BNP Paribas Asset Management (partie 4), roadmaps en phases documentées dans le cahier pratique DAMA France et ISACA AFAI (2021).
8. Pourquoi consulter IDAIA	Positionnement concurrentiel et architecture des 5 piliers IDAIA.